

人工知能 (AI) の学習用データに関する 知的財産の保護

会員・弁護士 岡本 義則

要 約

近年の人工知能 (AI) の技術, 特にニューラルネットワークにおける深層学習 (ディープラーニング) の技術の発展により, 人工知能 (AI) についての期待が高まっている。しかし, 深層学習については, 学習用のデータをどのように取得するかが問題となる。

一方, 知的財産の分野においては, 第 4 次産業革命に即した知的財産制度が模索されており, データ・人工知能 (AI) の利活用促進による産業競争力強化の基盤となる知的財産システムの構築が検討されている。

本稿では, 人工知能 (AI) の学習用データ, 特にニューラルネットワークの学習用データに関して, AI の時代にふさわしい知的財産制度に求められる機能を検討し, 検討した知的財産制度について, 仮想事例を想定し, 制度の改正に必要な視点について検討する。

目次

1. はじめに
 - (1) 人工知能 (AI) の技術の発展
 - (2) 人工知能 (AI) 学習用データの問題
2. 人工知能 (AI) 学習用データの知的財産による保護
 - (1) AI 学習用データの保護の必要性
 - (2) AI 学習用データの保護の基本的な視点
 - (3) AI 学習用データの知的財産権による保護の具体例
3. 仮想事例による知的財産制度の検討
 - (1) 仮想事例
 - (2) 検討
4. 制度の改正における議論について
5. おわりに

1. はじめに

(1) 人工知能 (AI) の技術の発展

近年の人工知能 (AI) の技術, 特にニューラルネットワークにおける深層学習 (ディープラーニング) の技術の発展により, 人工知能 (AI) についての期待が高まっており, 第 3 次 AI ブームを迎えている。

4 層以上の多層のニューラルネットワークは, たとえば, ネオコグニトロンなど, 1980 年代のいわゆる第 2 次 AI ブームの以前から存在していた⁽¹⁾。筆者は第 2 次 AI ブームの末期にニューラルネットワーク及び人工知能 (AI) の研究をした⁽²⁾⁽³⁾⁽⁴⁾。当時は多層のニューラルネットワークについては, 誤差逆伝播法 (バックプロパゲーション) での学習には困難があった。

しかし, 近年, 多層のニューラルネットワークにおいて, 深層学習 (ディープラーニング) の技術が発展し, 画像認識等において, 大きな性能向上が得られた⁽⁵⁾⁽⁶⁾。

深層学習の進歩は, 学習アルゴリズムの改良もさることながら, 限られた状況ではあるが学習用の大量の教師付データが得られたことや, 学習を可能にするコンピュータの処理能力の向上が重要であったと考えられる。

深層学習については, その将来性に期待がもたれている⁽⁷⁾。日本は, 製造業など裾野の広い産業を有し, 様々な分野で学習に用いるデータを生み出しうるため, 深層学習により日本の産業競争力の強化が期待されている⁽⁸⁾⁽⁹⁾。

(2) 人工知能 (AI) 学習用データの問題

深層学習により画像の認識等で大きなブレイクスルーがなされたが, AI 学習用のデータとして, 大量の教師付データが必要となる点が問題となる。

ゲームなど教師付データを大量に利用できる分野では, AI が人間の能力を超える場面が出てきているが, 多くの実社会での応用においては, 学習用の教師付データが十分に集まらないという問題がある。教師なし学習, 半教師あり学習, 強化学習など様々な機械学習の手法が研究されているが, 学習用データとしては, 現在のところ, 教師付データが最も強力と考えら

れる。

そこで、AI学習用データ、特に大量の教師付データを、どのように収集し、利用するかが問題となる。

2. 人工知能（AI）学習用データの知的財産による保護

（1）AI学習用データの保護の必要性

一方で、人工知能（AI）の利活用促進による産業競争力強化の観点から、その基盤となる知的財産システムの構築が問題となっている。

平成29年3月に発表された、知的財産戦略本部の新たな情報財検討委員会の「新たな情報財検討委員会報告書－データ・人工知能（AI）の利活用促進による産業競争力強化の基盤となる知財システムの構築に向けて－」⁽¹⁰⁾においては、(1) AI学習用データ、(2) AIのプログラム、(3) AIの学習済みモデル、(4) AI生成物について、知的財産権の問題が議論されている。人工知能（AI）に関する知的財産権の問題については、(1)～(4)は、それぞれ全く性格の異なる問題と考えられる。同報告書においては、(1) AI学習用データに関して、特許権類似の知的財産権による保護の具体的な提案はなされていない。なお、(3) AIの学習済みモデルについては、特許庁「IoT関連技術等に関する事例について」(特許・実用新案審査ハンドブックにおける平成29年3月22日の事例追加)の事例2-14において、特許による保護の対象となりうるということが明確になっている⁽¹¹⁾。

AI学習用データの保護については、必ずしも有用性の明らかではない多種多様なデータを含む、一般的なデータの保護の問題とは異なる問題と考えられる。また、本稿は、AI学習用データとしては、人間が手間をかけて作成した良質なデータを検討の対象としており、IoT (Internet of Things) において、センサー等から自動収集される大量データについては、本稿の検討の対象外である⁽¹²⁾。

AI学習用データについては、数多くの良質なデータが集まるか否かがAIの性能に直結する。そこで、データの収集と利用が重要な問題となると考えられる⁽¹³⁾⁽¹⁴⁾。

本稿では、AI学習用データ、特にニューラルネットワークの学習用データの知的財産による保護について検討する。

（2）AI学習用データの保護の基本的な視点

AI学習用データの保護については、①できる限り多くの良質な学習用データを集めるにはどうすればよいか、②それを広範な人々が利用できるようにするにはどうすればよいか、が基本的な視点となると考えられる。

AI学習用データの保護については、営業秘密による保護を中心とする方向も考えられる。しかし、営業秘密による保護を受けるためには、いわゆる営業秘密の3要件（秘密管理性、有用性、非公知性）を満たさなければならない（不正競争防止法2条6項）。

営業秘密による保護をAI学習用データの保護の中核に据えた場合、AI学習用データについては、基本的に秘密として管理し、公知とならないようにする方向となる。たとえば、ある会社が有している営業秘密としてのAI学習用データを、別の会社に供与する場合でも、秘密保持契約を締結して、厳重な秘密管理をすることが前提になり、一般社会が、当該AI学習用データを利用することはできないことになる。

しかし、AI学習用データについては、AIの性能に直結し、AI社会における基本的なインフラストラクチャーとなりうるため、①できる限り多くの良質な学習用データを集め、②それを広範な人々が利用できるようにすることが、重要と考えられる。

そうすると、AI学習用データについては、秘密として管理しなくても、公開されて公知となっても、AI学習用データを保護できるように、営業秘密による保護ではなく、知的財産権による保護が必要となると考えられる。

（3）AI学習用データの知的財産権による保護の具体例

AI学習用データについて、知的財産権による保護を与えることは、学習用データを作成するインセンティブを与え、①できる限り多くの良質な学習用データを集めることに大きく貢献しうる。

一方で、知的財産権を付与すると、広範な利用が妨げられるのではないかと懸念もありうる。そこで、できる限り多くの良質な学習用データを集めつつ、②それを広範な人々が利用できるようにするにはどうすればよいか問題となる。

この点については、AI学習用データについては、著作権又は著作権に類似の権利で保護することも考えら

れる。しかし、著作権法は、文化の発展に寄与することを目的とする法律であり(著作権法1条)、産業の発達に寄与することを目的とする法律(特許法1条)ではないため、保護期間が著作者の寿命に依存するなど、産業の振興を目的とした法体系にはなっていない。

産業の振興のためには、特許法が長い歴史を有している。そこで、特許権を参考にして、特許権に類似の権利を与えることが、人工知能(AI)に関係する産業の振興に資すると考えられる。

AI学習用データに与える特許権に類似の知的財産権については、特許権ではないことを強調するために、「AI学習用データ権」などと呼ぶことも考えられる。しかし、本稿では特許権とのアナロジーにより理解を促進するため、暫定的に「データ特許」(仮称)と呼ぶことにする。

そして、データ特許については、特許法68条の「業として」の要件や、特許法69条1項の「試験又は研究のため」の要件を参考にして、それよりも広い範囲で無償の利用を認める規定を置くことが考えられる。

たとえば、学術利用、試験・研究のための利用、非営利の利用、小規模な営利利用については無償とし、大規模な営利利用(たとえばAI事業の売上額が1億円以上)についてのみ有償とすることが考えられる。

そして、出願の際に、権利を取得した場合には一定のライセンス料率で第三者にライセンスをするという条項(以下、便宜的に「自動ライセンス条項」という。)への同意を出願人に求めることにし、大規模な営利利用についても、権利を利用した場合の料金が明確になるようにすることが考えられる。ライセンス料率については、FRAND宣言の場合のように具体的な料率がいくらかについて見解の相違が生ずることを防止するため、固定のライセンス料率を出願人が設定し、公開公報に明記することが考えられる。

データ特許の出願は、①データクレーム、②データ明細書、③データ要約書など、特許出願と類似の書類と、④実際のAI学習用データ(寄託・公開される)により行なうことが考えられる。もっとも、出願の奨励のため、①、②は、出願時に必ずしも求めないことが考えられる。

データクレームは、ニューラルネットワークにおける教師付データの場合には、たとえば、以下のような記載となる。

【請求項1】 入力次元を10000、出力次元を1とする教師付データであり、植物Aの葉の入力画像(100×100)から、植物Aが病気Bにかかっている確率を出力する寄託済みの教師付データ。

このように、データクレームは、ニューラルネットワークの入力、出力次元など、データの形式を特定し、寄託済みのデータと共に、権利の対象を特定することが考えられる。権利の対象は、技術思想ではなく、学習用データである点が、特許権とは異なる。

次に、データ明細書は、データに関する詳細な説明をした明細書である。

【データの名称】には、データの簡単な名称を記載する。この例では、「植物Aの病気Bの識別データ」などの名称とすることが考えられる。

【データの詳細な説明】には、学習用データの属する分野、従来の学習用データの状況、学習用データの適用範囲と汎化能力、学習用データを適用した効果などについて、詳細な情報を記載する。

また、データ要約書は、データの名称、データの形式、代表的な使用例などを記載する。データ要約書は、権利範囲に影響しないため、出願人が自由にデータの説明を記載することも想定されている。

データ特許については、出願のみをして出願奨励金を受け取るという利用の仕方もあるが、審査請求をして、特許庁による審査を受けることもできる。

データ特許の審査においては、進歩性は必要なく、たとえば、①新規性、②学習可能要件、③学習有効性要件、④記載要件等を審査することが考えられる。

①新規性は、クレーム及び出願時に寄託されたデータが、従来のデータベースに存在しないかどうかを審査する。なお、他人の作成したデータを無断で出願すると冒認出願となる。

②学習可能要件は、特許における実施可能要件に対応するものであり、データを実際にAIに学習させることができるか否かを審査する。寄託されたデータに欠落があるなどして、学習ができない場合には、この要件を満たさない。

審査の際には、特許庁に備え付けられた複数の種類の標準的なニューラルネットワーク等を学習に用いて学習可能要件を判定する。標準的なニューラルネットワークは、一般に公開し、出願人が学習可能要件を事前にチェックできるようにすることが考えられる。ま

た、出願人が特殊なニューラルネットワークを用いる場合には、そのニューラルネットワークも寄託することで、学習可能要件を満たすとすることが考えられる。

③学習有効性要件は、特許における産業上の利用可能性（特許法 29 条 1 項柱書）に対応するものであり、たとえば、データを実際に AI に学習させた場合に、一応の性能を満たすか否かを審査する。この要件については、ある程度緩やかに審査することが考えられる。たとえば、汎化能力が足りず、画像の誤識別があっても、ある程度の認識率があれば、他のデータとの組み合わせで改善しうるため、要件を満たすとすることが考えられる。

④記載要件については、データクレームとデータ明細書の記載が権利付与にふさわしいものになっているかを審査する。なお、データクレームとデータ明細書の補正は、寄託された AI 学習用データと整合する限り自由に行なえるようにし、記載要件についての拒絶理由通知については、解消することが容易となるように制度設計をすることが考えられる。

審査の結果、データ特許査定となる場合には、審査官は効率的な学習のための情報を提供するために、データ特許メモを作成することができる。

出願にあたっては、公衆からできる限り多くのデータを集めるために、出願料は無料ないし低額に設定することが考えられる。さらに、寄託された AI 学習用データについて、公開されて一定以上の利用実績（ダウンロード数など）がある場合には、出願奨励金が得られるようにすることが考えられる。また、出願人には、今後の当該データの取引の際に用いることができる出願番号が与えられる。

AI 学習用データをできる限り多く集めるために、出願は AI 学習用データとデータ要約書があれば可能とすることが考えられる。その後、審査請求をする場合には、弁理士に依頼するなどして、データクレームとデータ明細書を出願後に追加することも可能なようにする。

データ特許は、出願時に、第三者への自動ライセンス条項に出願人が同意することを通常の取扱いとすることが考えられる。出願人が、低廉な料率の自動ライセンス条項に同意した出願については、出願料、審査請求料、データ特許料は、無料ないしは奨励金が得られるようにすることが考えられる。

一方、出願人が、自動ライセンス条項に同意しない場合には、通常の出願料、審査請求料、データ特許料を請求することが考えられる。

なお、他社は、自動ライセンス条項に同意されていない出願については、当該 AI 学習用データをダウンロードして使用しなければよい。

よって、出願人が自動ライセンス条項に同意をしない場合は、他社からライセンス料を得ることは通常は期待できず、AI 学習用データを一般に公開しつつ自社で使用する場合など、例外的なケースに限られるように制度設計することが考えられる。

3. 仮想事例による知的財産制度の検討

以下、上記のような「データ特許」制度が実現した場合の仮想事例を念頭において、「データ特許」制度に要求される機能について検討する。

(1) 仮想事例

A 氏は、作物 X の栽培に従事している農家であるが、気象や雑草の種類などの栽培の際の色々な条件と、与える肥料の量などについては、30 年にわたって詳細に記録を付けており、長年の経験を有していた。

A 氏は、後継者不足に悩んでおり、作物 X の栽培の際の条件と、栽培の際の肥料の量などの組について、自らの長年の経験に基づく記録をまとめて、教師付データを作成した。

A 氏は、特許庁の書式例を参考にしてデータ要約書を作成し、当面、データクレームとデータ明細書はなしで、データ特許を出願した。

A 氏は、売上額が 1 億円以上の営利利用について低廉なライセンス料率を定める自動ライセンス条項に同意をしたので、出願料は無料であった。形式審査を通った後、出願は公開された。出願公開後に、データのダウンロードの実績が一定数を越えたため、A 氏は出願奨励金を得た。

A 氏のデータのダウンロード実績は伸び続けた。まず、新規に就農した若者が、A 氏のデータを無料でダウンロードした。新規に就農した若者は、A 氏のデータや他の農家の出願したデータを用いてニューラルネットワークの学習をし、さらに、その農地特有の事情等について追加の学習をして、その出力を参考にして農業を行なった。

また、多くの研究機関において、A 氏のデータは無

料でダウンロードされ、他のデータと共に、農業関係の深層学習を用いたニューラルネットワークの性能評価など、多くの学術論文等において使用された。

さらに、農業関係の多数の企業も、製品のプロトタイプを作る際に A 氏のデータを用いた。小規模な販売をした企業もあったが、試験的な販売であり、売上額が1億円以上ではなかったため、A 氏のデータを無料で使用できた。なお、A 氏のデータを含め、一か月間の定められた量を超える莫大なデータ量のダウンロードを行なった企業は、ダウンロード利用料を特許庁に支払った。

このように、非常に多くの人々が、A 氏のデータを無料でダウンロードして、恩恵を受けた。

A 氏は、ダウンロード数から、データの有用性に確信を持ち、本格的に商用利用される可能性を感じたので、弁理士にデータクレームとデータ明細書の起草と審査請求を依頼した。

データクレーム、データ明細書が登録されると、データのダウンロード数はさらに伸びていった。ダウンロード数がさらに一定数を超えたため、A 氏は追加の出願奨励金を得た。

A 氏の出願は、特許庁において審査され、「データ特許」が付与された。A 氏は、出願の際に、低廉なライセンス料率の自動ライセンス条項に同意をしていたので、審査請求料と特許料は無料となった。

そして、ある農業機械メーカーが、A 氏のデータの有用性を認め、他の約 1000 名の農家の出願したデータと併用して、農業を自動化する農業機械の人工知能（AI）の学習用データとして活用することになった。

A 氏のデータを活用した農業を自動化する農業機械の人工知能（AI）は、莫大な利益を生み出した。

これにより、A 氏は、多くのライセンス料を得た。また、A 氏は、社会的に有用なデータの作成者としての名誉を得た。

A 氏のデータの貢献により、過疎化していた農地には活気が戻っていき、多くの農作物の収穫が得られた。

（2）検討

A 氏の長年にわたる農業における経験は、人工知能（AI）の社会において大きな役割を果たした。仮想事例においては、データ特許の制度により、①できる限り多くの良質な学習用データを集め、②それを広範な

人々が利用できるようにすることが可能になった。

このように、通常ならば、わざわざ人手をかけてまで作成されないか、死蔵されてしまうデータを、AI 学習用データに関するデータ特許の制度によって、社会において収集し、活用することが重要となると思われる。

企業ばかりでなく、様々な分野の多くの人々も、有用なデータを潜在的に保有しているか、作成可能であると思われる。しかし、現在のところ、そのようなデータは十分に活用されていない。データの出願を奨励し、出願後の利用実績に応じて出願奨励金を支払うことで、社会から広く有用なデータを集めることが考えられる。

そして、出願人は、データの利用実績等を検討して、産業界で広く利用される良質なデータと考えられる場合には、データクレーム、データ明細書を起草し、審査請求をして、データ特許を取得することができる。もちろん、AI 学習用データの提供を業として行なう企業などは、当初からデータクレーム、データ明細書を起草して出願することが考えられる。

個人の提供したデータでも、多くの利用実績を積み重ね、その分野の有用な AI 学習用データの一つとなり、産業界に貢献すれば、奨励金ないしはライセンス料の取得が可能となるように制度を設計することが考えられる。

4. 制度の改正における議論について

データ特許（仮称）については、進歩性などの考え方にはなじまないため、通常の特許とは別の新しい知的財産権と考えられる。新しい知的財産権を創設する法律の改正が行なわれる場合、その内容の詳細については、審議会等での議論が行なわれると考えられる。その際には、一般からの意見が募集されると考えられる。

データ特許（仮称）については、従来にない制度の提案であり、法律の改正で採用する場合には、その具体的な内容については、産業界、AI 関係者、知的財産関係者だけではなく、データの提供主体となりうる人々を含めて、多くの意見が検討されることが必要と思われる。

その際には、AI 学習用データについては、①できる限り多くの良質な学習用データを集めるにはどうすればよいか、②それを広範な人々が利用できるようにす

るにはどうすればよいかという視点が重要となると考えられる。そのためには、現在は知的財産権の制度に関心がないが、このような制度があれば利用してみたいという層のニーズを取り込んでいくことが、AI学習用データの利用促進と産業の発展のために重要となると思われる。

5. おわりに

本稿では、人工知能 (AI) の学習用データに関して、今後の AI の時代にふさわしい知的財産制度に求められる機能を検討し、検討した知的財産制度について、仮想事例を想定して具体的に検討し、制度の改正に必要な視点について検討した。

AI 学習用データの保護については、不正競争防止法による保護を中心とし、秘密にする方向で法制化をすることも考えられる。しかし、AI の性能の向上のためには、AI 学習用データについては、①できる限り多くの良質な学習用データを集めること、②それを広範な人々が利用できるようにすること、が重要となると考えられる。

AI 学習用データについて、①と②を実現するには、新たな知的財産権の導入が最も適切な方法と考えられるが、その具体的な内容については今後の議論が必要であり、改正法によって導入するためには、多くの人々による議論が必要となると考えられる。

本稿が、AI 学習用データの知的財産制度による保護について、今後の議論の一助となれば幸いである。

(参考文献)

- (1) 福島邦彦「位置ずれに影響されないパターン認識機構の神経回路モデル—ネオコグニトロン」, 電子情報通信学会論文誌 Vol.J62-A, No.10, pp.658-665 (1979)
- (2) 岡本義則「環境の激しい変化に適應する神経回路網モデル」, 電子情報通信学会論文誌 Vol.J73-D- II ,No.8, pp.1186-

- 1191 (1990)
- (3) 岡本義則, 中島秀之, 大澤一郎「確信度と主観確率を持つ信念推論システム」, 人工知能学会論文誌 Vol.7, No.2, pp.263-270 (1992)
- (4) 岡本義則「定量的物理モデルを用いた幾何学的推論」, 電子情報通信学会論文誌 Vol.J75-D- II , No.11, pp.1866-1873 (1992)
- (5) Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton "ImageNet Classification with Deep Convolutional Neural Networks", Advances in Neural Information Processing Systems 25, NIPS (2012).
- (6) 岡谷貴之「ディープラーニングによる画像認識 - 畳込みネットワークの能力と限界 -」, 情報処理 Vol. 56, No. 7, pp. 634-639 (2015)
- (7) 松尾豊「人工知能の未来: ディープラーニングの先にあるもの」, 技術と経済 No. 595, pp.10-25 (2016)
- (8) 大堀達也, 池田正史「深層学習で激変するビジネス『AI 大国』へ正念場の日本」, エコノミスト Vol. 94, No. 20, pp. 18-21 (2016)
- (9) インタビュー 松尾豊氏 東京大学 大学院工学系研究科技術経営戦略学専攻 特任准教授「深層学習の価値は『目』の獲得産業応用で日本は勝てる」, 日経コンピュータ No. 929, pp. 48-51 (2017)
- (10) 知的財産戦略本部 検証・評価・企画委員会 新たな情報財検討委員会「新たな情報財検討委員会報告書 - データ・人工知能 (AI) の利活用促進による産業競争力強化の基盤となる知財システムの構築に向けて -」, 平成 29 年 3 月
- (11) 特許庁「IoT 関連技術等に関する事例について」(特許・実用新案審査ハンドブックにおける 2017 年 3 月 22 日の事例追加), pp. 41-44 (2017)
- (12) 上野達弘「自動集積される大量データの法的保護」, パテント, Vol.70, No.2, pp.30-36 (2017)
- (13) 小田桐優理「コンテンツプラットフォームにおける機械学習, データセット公開・モデル公開による産学の発展」電子情報通信学会誌 Vol. 100, No. 1, pp. 25-31 (2017)
- (14) インタビュー 辻井潤一氏 産業技術総合研究所 人工知能研究センター 研究センター長「米国と異なる AI 研究体制目指す データを持つ企業との連携が鍵」, 日経コンピュータ No. 911, pp. 36-39 (2016)

(原稿受領 2017. 5. 9)